面向少样本网状结构体的候选区域自适应检测方法 *

牟　磊[a, b]，陈　黎[a, b]

(武汉科技大学 a. 计算机科学与技术学院; b. 智能信息处理与实时工业系统湖北省重点实验室，武汉 430065)

摘　要：在仅有少量标签数据的图像网状结构检测任务中，需要大量训练数据的目标检测模型检测性能大幅下降。基于区域候选的目标检测模型在预测时，检测目标越多，检测时间越长。若基于区域候选的目标检测模型产生候选框的数量固定不变，而不同图像中网状结构目标数量不同，造成目标检测中额外的时间消耗。针对该问题，通过对训练样本中网状结构目标在图片中的密度分析以及根据网状结构体在图片中的特征分布，提出一种面向少样本网状结构体的候选区域自适应检测方法。该方法通过基于二值标签图标注方法得到大量训练样本，由候选区域自适应方法选取合理候选框数量。与未改进的模型相比，在几乎不损失准确率的情况下，加快了检测速度，尤其在目标数量稀少的数据中优势更为明显。

关键词：少样本；网状结构体；样本标注；候选区域自适应

中图分类号：TP391.4　　doi: 10.3969/j.issn.1001-3695.2018.07.0427

Proposal adaptive detection method for small sample reticular structures

Lei Mou[a, b], Li Chen[a, b]

(*a. School of Computer Science & Technology, b. Hubei Province Key Laboratory of Intelligent Information Processing & Real-time Industrial System, Wuhan University of Science & Technology, Wuhan* 4300653, *China*)

Abstract: The detection performance of the object detection model requiring a large amount of training data is greatly reduced in the reticular structures detection task with a small amount of labeled data. The detection model which based on region proposals will spend longer time with larger outputs during predicting. And the number of reticular structures are different in different images. It will result in extra time consumption in object detection if the number of proposals is fixed in different images. In view of this problem, we propose an proposal adaptive detection method for small sample reticular structures by analyzing the density of the reticular structures in the training samples and according to the characteristic distribution of the reticular structures in the images. In this method, a large number of training samples are obtained by the method of binary-value labeled map marking, and a reasonable number of proposals are selected by the proposal adaptive method. Compared with the unimproved model, the detection speed is accelerated without loss of accuracy, especially in the data with few objects.

Key words: small sample; reticular structure; sample labeling; proposal adaptive

0　引言

深度学习在视觉识别任务中取得了巨大成功[1,2,3]，然而，这些模型需要大量标签数据和迭代次数来训练大量参数。实际中会面临少量甚至没有标签数据的情况，且人工标注标签数据工作繁重，代价太大。目前，针对标签数据少的问题，涌现了许多诸如数据增强、少样本学习[4]等技术。

目前，深度学习目标检测方法主要分为两类：基于区域候选的模型和基于回归的模型[5]。基于区域候选的深度学习目标检测模型建立在区域候选的思想上，首先对检测区域提取候选框，然后再将提取的候选框用于后续特征提取、候选框回归和分类[6]。在基于区域候选的模型中，比较典型的方法包括R-CNN[7]、Fast R-CNN[8]、Faster R-CNN[9]等一系列方法。基于回归的模型采用回归的思想，需要预先按照一定方式划定默认框，从而建立起预测框、默认框、ground-truth 物体框的关系以进行训练[20,21]，如 YOLO[10]，SDD[11]等一些列优秀模型。对于第一类方法，目标检测性能表现优秀，但相对第二类方法其检测速度很慢；相对于第一类方法，第二类方法检测性能下降，

收稿日期：2018-07-10；**修回日期：**2018-08-28　　**基金项目：**国家自然科学基金资助项目（61773297，61375017）；智能信息处理与实时工业系统湖北省重点实验室开放基金资助项目（2016znss01A）

作者简介：牟磊（1994-），男，湖北恩施人，硕士研究生，主要研究方向为图像处理、深度学习（1144545359@qq.com）；陈黎（1977-），男，教授，博导，主要研究方向为智能媒体计算与安防视频应用、图像处理、深度学习.

其检测速度提升明显 [1,2]。

基于区域候选的模型通过区域候选网络产生候选框，再将候选框输入到全连接层进行多分类和回归计算。由于全连接层参数众多，在预测时每一个神经元都参与运算，若全连接层的输入增多，则预测时间成倍增加[12]。不同网状结构体样本，待检测目标数量不同；对于不同数据，在区域候选阶段，以固定不变的输出作为全连接层的输入，使得参与全连接层计算的候选框多于待检测目标，造成额外时间耗费。同时在纵横交错、相互连接以及呈树形拓扑分布的网状结构体标注过程中，候选框标签需连续分布以覆盖整个网状结构体，导致人工标注工作繁重，耗时耗力，且人工标注不精确等问题。

针对以上两方面问题，本文提出一种面向少样本网状结构体的候选区域自适应检测方法，该方法通过两级预处理将网状结构体划分成线状集合，再将线状集合划分为块状集合，最后获得大量精确标注的候选框，达到数据扩增的效果。同时，该方法根据不同训练数据中目标密度，设计一种候选区域自适应算法，选取合适候选框数量作为区域候选网络的输出，从而减少全连接层的计算耗时，达到在几乎不降低模型性能的前提下，加快预测速度的效果。实验结果表明，该方法对面向小样本网状结构体检测有着良好的性能表现，对不同数据的检测时间有着明显改善。

1　国内外研究现状

1.1　数据增广

深度学习模型的训练需要大量标签数据，现实中，当只有少量标签数据时，训练的深度学习模型性能大幅下降。数据增广技术为面临少量标签数据训练模型的情况提供了一种解决方案。数据增广技术指在不改变图像类别的情况下，增加数据量，提高模型的泛化能力。从几何角度看，常用的数据增广技术包括翻转，位移，裁剪等；从图像像素变换角度看，常用方法包括颜色抖动，增加噪声等。

1.2　少样本学习

近年来，针对标签数据量少的情况，国内外进行了大量的研究。早期方法将在源域(source domain)数据中学习的经验知识用于目标域(target domain)训练和学习的迁移学习技术(transfer learning)[13]为少样本学习开拓了新道路。迁移学习技术对深度学习模型性能提升明显，但模型性能的提升仍然具有很大空间。目前深度学习方法在训练过程中，利用神经网络提取目标特征，然后对特征进行分类以及回归，然而在这一过程中，大多数方法将焦点放在样本独立性上，没有考虑同类样本之间以及目标之间的相似关系，基于此，出现了孪生网络(siamese neural network) [14]，匹配网络(matching network) [15]，原型网络(prototypical networks) [16]，关系网络(relation network) [17]等一系列新颖网络，这些网络将焦点集中于学习样本之间的相似关系，使得模型拥有比较样本之间相似度的能力，从而利用少量标签样本，通过相似度比较，实现分类和识别的目的。

1.3　基于区域候选的模型

目前，目标检测模型主要包含两类：基于区域候选的模型；基于回归的模型。基于区域候选的模型是一种两次检测(two-stage)深度神经网络，而基于回归的模型是一种单次检测(one-stage)深度神经网络。相较于单次检测神经网络，两次检测神经网络的精度更高。早期方法如 R-CNN[7]、Fast R-CNN[8]使得目标检测取得了长远的进步，Faster R-CNN[9]的出现进一步提高了基于区域候选模型的检测速度和检测精度。对区域候选网络从 Selective Search[7]到 RPN[9]的改进加快了模型预测速度，在区域候选网络之后加入全连接层对候选区域分类回归提高了模型预测精度。

1.4　网状结构体

现实生活中，网状结构体无处不在。对人体静脉血管，视网膜眼底血管等网状结构体的检测与分割有助于及时发现病患，降低人体健康危害；对卫星道路，地面裂缝等网状结构体的检测有助于降低人工检测的安全风险[22]。目前，针对网状结构体的分割方法主要基于 U-Net[18]展开研究，而针对网状结构体的检测方法相对较少，由于网状结构体结构复杂，对模型检测精度要求高，因此对网状结构体的检测主要基于区域候选模型展开研究。

2　面向少样本网状结构体的候选区域自适应检测

2.1　面向少样本网状结构体检测

在网状结构体二维图像中，有许多交错并相互连接的线状集合。例如图 1(a)所示视网膜血管中以视盘为中心呈树根状分布的网状结构体，图 1(b)所示的道路图像中呈狭长面条状的网状结构体。

网状结构体图像的网状结构纵横交错、相互连接且分布广泛。康定斯基理论[19]认为，点、线、面构成平面空间的基本元素。受此理论启发，将网状结构体看作二维图像中的“面”，进一步将网状结构体进行第一层划分，使之成为线状结构体，将线状体看作“线”，再对线状体进行第二层划分，使之成为块状体，将块状体看作“点”。通过对网状结构体的分层处理思想，采用基于二值标签图标注方法对样本进行标注，最终获得用于训练的候选框。整个分层处理过程如图 2 所示。图 2 中，在左侧虚线区域内黄色边框代表第一层分解得到线状体结合，绿色边框代表第二层分解得到块状集合，得到块状体后，根据步骤 d)计算 IoU，生成候选框，最后根据二值标签图生成的候选框坐标，在训练样本中进行标注。



(a)呈树型拓扑结构分布的视网膜血管网状结构体



(b)纵横交错、相互连接的卫星道路网状结构体

图 1　网状结构体示意图

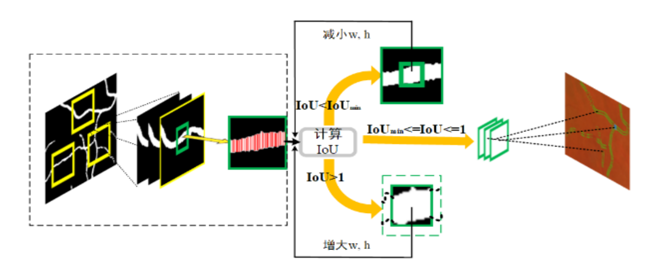


图 2 基于二值标签图标注方法流程图

基于二值标签图标注方法步骤如下：

a)根据二值化标签图(图 3(a))，对前景区域进行骨架提取，然后以骨架坐标作为候选框的中心坐标，选取第一个候选框；

b)以当前候选框边长 $\frac{1}{m}$ 的距离，在骨架坐标上重新找点，作为下一个候选框的中心坐标；

c)以当前候选框边长 w,h 以及第 b)步生成的候选中心生成新的候选框；

d)计算被新候选框覆盖的标签 ground_turth 面积与新生成的候选框面积的比值(如图 3(b))，即

$$IoU=\frac{S_{ground_truth}}{S_{new_proposal}}$$

e)当 $IoU\in[IoU_{\min},1]$ 时，则完成新候选框的生成，然后执行 b)~e)继续生成候选框；

f)若 $IoU<IoU_{\min}$，则新候选框中心坐标不变，以当前候选框边长 m 倍为新候选框边长，重新生成候选框，重复步骤 d)e)直至满足 $IoU\in[IoU_{\min},1]$；

g)若 $IoU>1$，则以 $\frac{1}{m}$ 替换步骤 f)中 m，继续步骤 f)

h)重复步骤 b)~g)直至结束。

采用基于二值标签图标注方法，第一层分解将网状结构体分成含有线状集合的 N_1 个部分，第二层分解将线状集合划分成含有块状体集合的 N_2 个块，最后对块状体进行标注；若网状结构体样本总数为 N_0，则分级算法生成的候选框数量为

$$N_{proposal}=N_0\times N_1\times N_2 \tag{1}$$

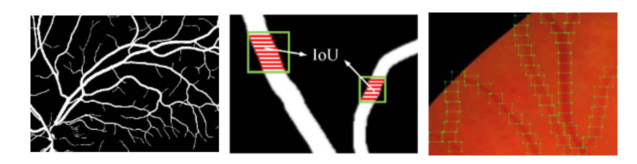
另外，在算法标注过程中，生成新候选框的中心坐标与当前后候选框边长有关，即

$$(x_{i+1},y_{i+1})=\left\{(x_{i+1},y_{i+1})\middle|\sqrt{(x-x_i)^2+(y-y_i)^2}=\frac{1}{m}(\min\{w_i,h_i\})\right\} \tag{2}$$

其中：(x_{i+1},y_{i+1}) 表示新候选框的中心点坐标，(x,y) 表示骨架线上的坐标点，(x_i,y_i) 表示当前候选框中心坐标，w_i,h_i 表示当前候选框长和宽。设当前候选框为 $P_i(x,y,w,h)$，则根据基于二值标签图标注方法可求得每一次迭代中，新候选框为

$$P_{i+1}=C_{i+1}(P_i)+\lambda\cdot L_{IoU}(w_{i+1},h_{i+1}) \tag{3}$$

式(3)表示新生成候选框的中心点坐标需要根据当前候选框决定，同时新候选框的大小和新的 IoU 有关，其中 $C_{i+1}(P_i)$ 表示以当前候选框生成新候选框的中心坐标，$L_{IoU}(\bullet)$ 表示根据 IoU 值计算新候选框的边长，基于二值标签图标注方法最终效果如图 3(c)所示。



(a) 基于二值标签图标注方法在标注时所需的二值化标签图 (b) 基于二值标签图标注方法中标注候选框时 IoU 示意图,红色阴影为 IoU,定义为 ground_turth 面积与候选框面积的比值 (c) 通过基于二值标签图标注方法进行标注的结果图,其中绿色边框表示标注的候选框,候选框标注精确且连续分布

图 3 二值标签图标注方法

根据式(1)，当网状结构体样本较少时，可通过基于二值标签图标注方法，标注大量候选框，扩充训练样本，为网状结构体训练样本量少提供了解决方案。

2.2 候选区域自适应

在生成候选框阶段，在一张图像中随机生成 $N_1\times N_2$ 个锚点(anchors)，然后对每一个锚点进行分类和回归，最终输出前 N 个候选框(proposals)。在 Faster R-CNN[15]中，其损失函数可简化为

$$L(p_i,t_i)=L_{cls}(p_i,p_i^*)+\lambda p_i^* L_{reg}(t_i,t_i^*) \tag{4}$$

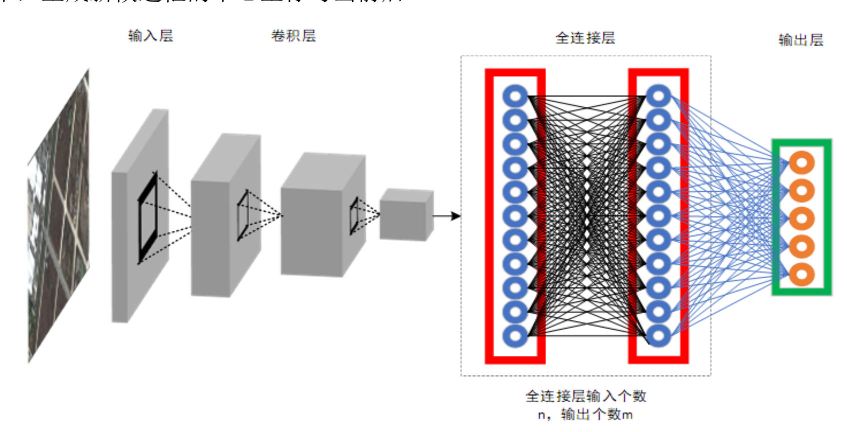


图 4 全连接层示意图。绿色表示全连接层的输入，$x_1,x_2,\cdots,x_n$ 表示全连接层神经元，$y_1,\cdots,y_m$ 表示全连接层输出结果

式(4)与文献[15]中不同之处在于，公式(4)未对锚点(anchors)坐标进行归一化，其中 i 表示锚点的编号，p_i 表示第 i 个锚点预测为目标的概率，p_i^* 表示第 i 个 ground-truth 标签值，当第 i 个锚点判定为正样本时，$p_i^*=1$，当第 i 个锚点判定为

负样本时，$p_i^*=0$，t_i 与 t_i^* 与第 i 个锚点的坐标相关。从公式(4)可以看出，在最优化过程中，需要对每张图片中生成的每个锚点进行分类与回归优化，最终优化出前 N 个候选框作为该阶段的输出。

经过候选生成阶段产生的候选框特征，继续输入全连接层(fully connected layers, FC)进行分类和回归，在该阶段中，全连接层在神经网络中起到分类器的作用，将学到的分布式特征表示映射到样本标记空间[20]。如图 4 所示为一个全连接层示意图，区域候选网络产生 N 个 $n\times n$ 大小的候选框，则全连接层输入大小为 $n\times n\times N$，且全连接层的输出大小为 k。因此全连接层的计算大小为 $n\times n\times N\times k$，其计算公式为

$$(x_1 \quad x_2\cdots x_n)\times\begin{pmatrix} w_{11} & \cdots & w_{m1} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{mn} \end{pmatrix}+(b_1 \quad b_2\cdots b_m)=(y_1 \quad y_2\cdots y_m) \tag{5}$$

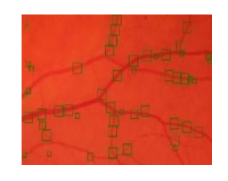
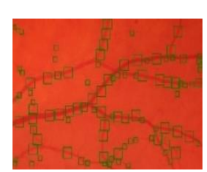
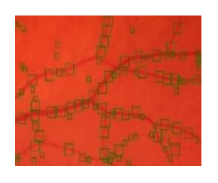
其中：w 为权重矩阵，b 为偏置。因此在全连接层需要对每一个神经元进行计算，才能得到最终的结果。若计算一个神经元的时间消耗为 t，则计算全部神经元的时间消耗为

$$T(N)=7\times 7\times N\times kt \tag{6}$$

其时间消耗与候选区域 N 呈线性关系。

通过式(6)及以上分析可得，在候选框生成阶段，产生的候选区域 N 越小，模型预测时间越短。因此可通过减少候选区域 N 达到加快模型预测速度的目的。但通过多次试验发现，当 N 减小到一定值后，目标检测模型出现大量漏检现象。如图 5 所示。其中(a)~(c)分别表示候选框设定为 50、100、300 时，最终检测出网状结构体中的目标个数，(a)出现大量漏检现象，(b)和(c)表现出在 N=100 和 N=300 时，检测出的目标个数都为 98。

为解决生成候区域阶段对不同数据因输出候选框个数 N 过大造成额外预测时间消耗，以及因候选区域数量设置过小出现大量漏检的问题，本文提出候选区域自适应算法，针对不同的数据，自适应选取合理候选区域数量。

(a)N=50,检测结果=50 (b)N=100,检测结果=98 (c)N=300,检测结果=98

图 5 同一张图片在不同 N 值下的预测情况

在 2.1 节中,每一张图片经由基于二值标签图标注方法生成许多候选框,每一张图片中块密度定义为图片中标注候选框数量与图片面积的比值,即

$$\rho_i=\frac{N_{proposal}}{S_{img}} \tag{7}$$

即图片中单位面积内候选框的个数。很显然，图像的块密度越大，代表着图像中的目标越多，所以可以根据块密度 ρ_i 来计算目标数量。由于训练样本中不同图片有着不同的块密度，并且区域候选网络接受任意大小图片，同时又要保证模型能够检测出所有目标，故采用训练样本中块密度最大值 $\rho_{\max}$，即

$$\rho_{\max}=\max\{\rho_i | i\in[1,\mathrm{N}]\} \tag{8}$$

来计算目标数量。

在模型预测时，根据测试图片的长(H)、宽(W)，利用式(7)计算出当前测试图片在候选生成阶段的 N 值，即

$$N=\rho_{\max}\times W\times H \tag{9}$$

以此计算出最终的候选区域，既能满足几乎不降低模型性能，又能达到加快模型预测速度目的。另外，当块密度很小时，自适应算法计算出的 N 可能会不足以满足真实目标个数值，并且在目标数量很小时，模型对 N 值是敏感的，为防止计算出的 N，在目标数量很少的情况下出现偏差，遂将 N 最小值设定为 M。因此，候选区域自适应算法最终产生候选框个数为

$$N=\max\{M,\rho_{\max}\times W\times H\} \tag{10}$$

表 1 两种模型在准确率和时间上的对比

(测试平台:CPU Premium G4560. 8GB DDR4, Windows 10)

数据	方法	N	mAP(%)	时间(s)
STARE	SSD	-	19.4	2.140
	YOLO-v3	-	21.7	1.451
	Faster R-CNN	300	74.8	17.937
	候选区域自适应	156	74.8	9.502
卫星道路	SSD	-	30.6	2.203
	YOLO-v3	-	47.2	1.238
	Faster R-CNN	300	93.6	17.253
	候选区域自适应	50	93.4	5.598

3 实验结果与分析

3.1 实验数据

网状结构体检测模型，检测出的结果是许多块状集合，每一个块状体，包含了自身在图像中位置信息以及所属类别。文中的方法在于，利用块状结构反映出网状结构的分布，因此，对块状体的分类正确，定位准确是最期望的结果。为了测试模型对网状结构体的检测性能，利用 2 组不同的数据集进行测试，计算每组数据中的准确率。第一组数据为 STARE[22]眼底图片库，训练样本为分辨率 3504×2366 的图片，在实验过程中由于分辨率太大，所以将一张图片九等分成分辨率为 1168×779 的九张图片，最终训练样本为 90 张图片，测试样本为 36 张图片；第二组数据为卫星道路，训练样本为分辨率 1862×900 的 12 张图片，测试样本为 12 张图片。

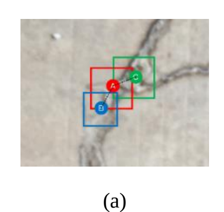
3.2 平均精度均值

在目标检测领域，检测结果的衡量通常采用平均精度均值(mean average precision, mAP)作为检测结果的准确性评判标准。对于检测目标中的每一个类别 C，可其精确率(Precision)与召回率(Recall)绘制一条曲线。曲线包围面积表示平均精度(AP)，求得多个类别的平均精度的平均值，得到平均精度均值 mAP。mAP 值越大，表示模型检测结果越准确，即图像中网状结构目标检测越完整。候选区域自适应检测方法检测结果如图

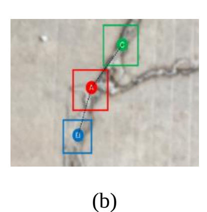
3-2 所示。通过模型检测出的图像中网状结构，覆盖率分别达到 74.8%(STARE)与 93.4%(卫星道路)，很好的反映了图像中网状结构的基本构成分布特征，为后期视网膜血管分割以及道路网的检测提供有力支撑。

3.3　样本扩增

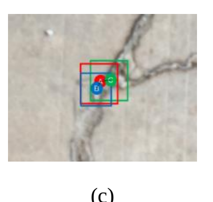
利用基于二值标签图标注方法产生大量训练样本，为少样本情况下网状结构体的检测提供了解决方案。在基于二值标签图标注方法中，生成新候选框中心坐标时，通过式(2)，以当前候选框短边长的 $\frac{1}{m}$ 倍距离，在中心线坐标中寻找满足要求的坐标点。如图 6 所示，A,B,C 分别代表三个不同候选框的中心点坐标，黑色虚线表示中心坐标之间的距离，设三个候选框的短边长分别为 l_A,l_B,l_C，由基于二值标签图标注方法可知，A, B 两中心坐标之间的距离为 $\Delta l=\frac{1}{m}l_A$，当 $\frac{1}{m}=\frac{1}{2}$ 时，B 点始终在候选框 A 的范围内，使得分级算法标注过程中不会出现因两个候选框重叠区域太小导致部分目标未标注的情况，如图 3-1(b)所示；同时不会出现因候选框重叠区域太大导致目标重复标注的情况，如图 3-1(c)所示。通过样本扩增，分别获得两组数据最终标注候选区域约为 14000、600。



(a)



(b)



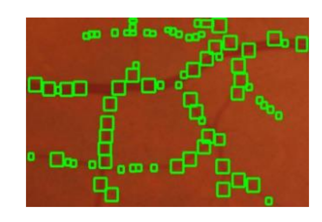
(c)

图 6　(a)当选取 $\frac{1}{m}=\frac{1}{2}$ 时,新候选框中心坐标始终在当前候选框范围内。(b)当 $\frac{1}{m}>\frac{1}{2}$,新候选框中心坐标距离当前候选框中心坐标过大,造成候选框标注不连续,部分网状结构体漏标注。(c)当 $\frac{1}{m}<\frac{1}{2}$ 时,新候选框中心坐标距离当前候选框中心坐标过小,造成候选框重复标注

3.4　自适应选取候选区域

通过两组数据测试区域候选自适应算法，分别计算比较区域候选自适应检测模型与常规目标检测模型在同一组数据下的预测准确率(mAP)以及时间消耗。结果如表 1 所示。









原始图片　　　　预测结果图

图 7　基于候选区域自适应检测方法对 STARE 数据集以及卫星道路数据集的检测结果图

从表 1 可知，通过候选区域自适应方法大幅减少区域候选网络的输出后，与基于回归的检测模型 YOLO-v3、SSD 相比，候选区域自适应检测方法的检测结果准确率提高约 43%；与基于区域候选的模型 Faster R-CNN 相比，在没有损失模型检测准确度的前提下，大大提高了模型检测速度，时间减少约 57%。由于候选区域自适应方法属于基于区域候选的模型，在模型中加入了全连接层进行分类和回归，而 YOLO、SSD 等基于回归的模型，未加入全连接层进行分类回归。但本文提出的方法在预测速度上接近 YOLO 和 SSD，同时模型预测准确度远远高于 YOLO、SSD 等一系列方法。

为继续探究区域候选自适应方法对模型性能的影响，以 STARE 和卫星道路数据为研究对象，分别以 N 从 1~400，计算候选区域 N 与模型性能以及预测时间的关系，其结果如图 8、9 所示。图 8 表示不同候选区域数量下模型性能表现曲线，横轴表示候选区域数量，纵轴表示检测精度度量标准 mAP，其中两个绿色菱形点代表由区域候选自适应方法产生的候选区域下，模型的预测性能(mAP)；图 9 表示不同候选区域数量下模型预测时间曲线图，横轴表示候选框数量，纵轴表示检测时间，其中两个绿色菱形点表示在由区域候选自适应方法产生的候选区域下，模型的预测时间。

由图 8 可知，候选区域个数从少到多的过程中，模型性能逐渐收敛到最大值，在 STARE 和卫星道路数据中，当 N 分别达到 159 和 50 时，模型的性能不再提高；由图 9 可知，模型预测时间与候选区域个数大体上呈线性关系，候选区域个数增加，模型预测时间变长。结合图 8、9 可知，通过候选区域自适应方法计算出区域候选网络输出候选框的个数值，恰好处于在模型性能充分发挥的情况下，预测速度最快的临界点；即当候选区域数量大于临界值时，若继续增加候选区域个数，模型性能不再提高，反而降低了模型检测速度，当候选区域数量小于临界值时，模型性能大幅下降。

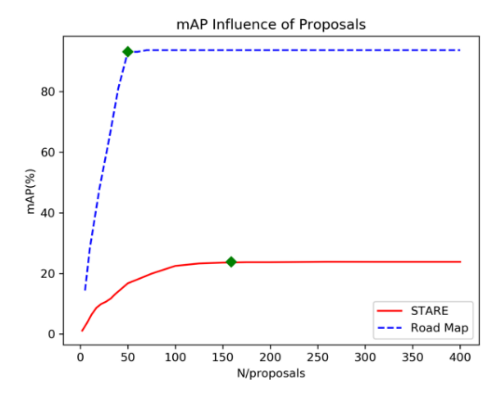


图 8　不同候选区域个数下模型准确率关系图

4　结束语

本文通过分析候选区域的生成方式以及全连接层的计算方式，提出一种候选区域自适应方法，实现候选区域自适应选取的目的，从而达到在不降低模型性能的前提下加快模型预测速度的效果；同时根据网状结构体的分布特征与结构特征，提出

基于二值化标签标注方法，在少量网状结构体标签样本情况下，获得大量可训练候选区域，达到数据扩增的效果。综合以上两个方面，本文提出的面向少样本网状结构体的候选区域自适应检测方法，通过 STARE 和卫星道路数据，验证了本方法在少样本网状结构体情形下，在不降低模型性能的同时，加快了模型预测速度。

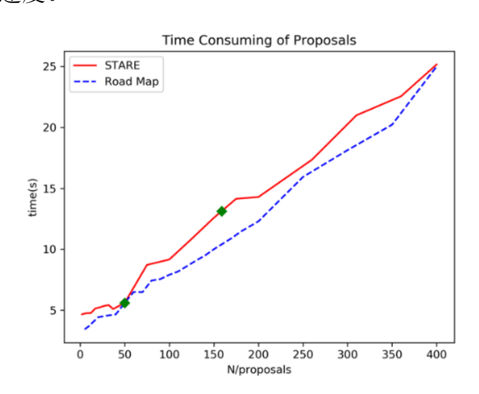


图 9　不同候选区域个数下模型预测时间图

参考文献：

[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. Curran Associates Inc. , 2012: 1097-1105.

[2] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014.

[4] Xian Yongqin, Lampert C H, Schiele B, *et al.* Zero-shot learning: a comprehensive evaluation of the good, the bad and the ugly [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2018.

[5] 吴帅，徐勇，赵东宁. 基于深度卷积网络的目标检测综述 [J]. 模式识别与人工智能，2018, 31 (4): 335-346. (Wu Shuai, Xu Yong, Zhao Dongning. Survey of object detection based on deep convolutional network [J]. Pattern Recognition and Artificial Intelligence, 2018, 31 (4): 335-346.)

[6] 李旭冬，叶茂，李涛. 基于卷积神经网络的目标检测研究综述 [J]. 计算机应用研究，2017, 34 (10): 2881-2886. (Li Xudong, Ye Mao, Li Tao. Review of object detection based on convolutional neural networks [J]. Application Research of Computers, 2017, 34 (10): 2881-2886)

[7] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.

[8] Girshick R. Fast R-CNN [C]// Proc of IEEE International Conference on Computer Vision. 2015: 1440-1448.

[9] Ren Shaoqing, He Kaiming, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [C]// Advances in Neural Information Processing Systems. 2015: 91-99.

[10] Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779-788.

[11] Liu Wei, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector [C]// Proc of European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[12] Chen Liangchieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS [J]. IEEE Trans on Pattern Analysis and Machine Intelligence 2018, 40 (4): 834-848.

[13] Pan J S, Yang Qiang. A survey on transfer learning [J]. IEEE Trans on knowledge and data engineering, 2010, 2 (10): 1345-1359.

[14] Gregory K, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition [C]// Proc of ICML Deep Learning Workshop. 2015.

[15] Oriol V, Blundell C, Lillicrap T, *et al.* Matching networks for one shot learning [C]// Advances in Neural Information Processing Systems. 2016: 3630-3638.

[16] Touretzky D S, Mozer M C, Hasselmo M E. Advances in neural information processing systems [C]// Cambridge: MIT Press, 1996.

[17] Sung F, Yang Yongxin, Zhang Li, *et al.* Learning to compare: relation network for few-shot learning [EB/OL]. (2018-03-27) . https://arxiv. org/abs/1711. 06025.

[18] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [C]// Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.

[19] Kandinsky, Wassily, Hilla Rebay. Dover Publications. Point and line to plane [M]. Courier Corporation, 1979.

[20] Viola P. Robust real-time object detection [C]// Proc of International Workshop on Statistical and Computational Theories of Vision: Modeling, Learning, Computing, and Sampling. 2013: 87.

[21] 周飞燕，金林鹏，董军. 卷积神经网络研究综述 [J]. 计算机学报，2017, 40 (6): 1229-1251. (Zhou Feiyan, JIN Linpeng, Dong Jun. Review of Convolutional Neural Network [J]. Chinese Journal of Computers, 2017, 40 (6): 1229-1251)

[22] 王海，蔡英凤，贾允毅，等. 基于深度卷积神经网络的场景自适应道路分割算法 [J]. 电子与信息学报，2017, 39 (2): 263-26. (Wang Hai, Cai Yingfeng, Jia Yunyi, etal. Scene Adaptive Road Segmentation Algorithm Based on Deep Convolutional Neural Network [J]. Journal of Electronics & Information Technology, 2017, 39 (2): 263-26)